

# On Variable Density Compressive Sampling

Gilles Puy, Pierre Vanderghenst, and Yves Wiaux

**Abstract**—Incoherence between sparsity basis and sensing basis is an essential concept for compressive sampling. In this context, we advocate a coherence-driven optimization procedure for variable density sampling. The associated minimization problem is solved by use of convex optimization algorithms. We also propose a refinement of our technique when prior information is available on the signal support in the sparsity basis. The effectiveness of the method is confirmed by numerical experiments. Our results also provide a theoretical underpinning to state-of-the-art variable density Fourier sampling procedures used in MRI.

**Index Terms**—compressed sensing, variable density sampling, magnetic resonance imaging.

## I. INTRODUCTION

Compressed sensing demonstrates that sparse signals can be sampled through linear and non-adaptive measurements at a sub-Nyquist rate, and still accurately recovered by means of non-linear iterative algorithms. The theory requires incoherence between the sensing and sparsity bases and a lot of work has thus been dedicated to design such sensing systems [1].

In the present work, we concentrate on  $s$ -sparse digital signals  $\alpha = (\alpha_i)_{1 \leq i \leq N} \in \mathbb{C}^N$  in an orthonormal basis  $\Psi = (\psi_1, \dots, \psi_N) \in \mathbb{C}^{N \times N}$ . The vector  $\alpha$  contains  $s$  non-zero entries and its support is defined as  $S = \{i : |\alpha_i| > 0, 1 \leq i \leq N\}$ . We denote  $\alpha_S \in \mathbb{C}^s$  the vector made of the  $s$  non-zero entries of  $\alpha$ . This signal is probed by projection onto  $m$  vectors of another orthonormal basis  $\Phi = (\phi_1, \dots, \phi_m) \in \mathbb{C}^{m \times N}$ . The indices of the selected vectors are denoted  $\Omega = \{l_1, \dots, l_m\}$  and  $\Phi_\Omega^\dagger$  is the  $m \times N$  matrix made of the selected rows of  $\Phi^\dagger$ , where the symbol  $\cdot^\dagger$  stands for the conjugate transpose operation. The measurement vector  $y \in \mathbb{C}^m$  thus reads as

$$y = A_\Omega \alpha \text{ with } A_\Omega = \Phi_\Omega^\dagger \Psi \in \mathbb{C}^{m \times N}. \quad (1)$$

We also denote  $A = \Phi^\dagger \Psi \in \mathbb{C}^{N \times N}$ . Finally, we aim at recovering  $\alpha$  by solving the  $\ell_1$ -minimization<sup>1</sup> problem

$$\hat{\alpha} = \arg \min_{\alpha \in \mathbb{C}^N} \|\alpha\|_1 \text{ subject to } y = A_\Omega \alpha. \quad (2)$$

Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

G. Puy, P. Vanderghenst, and Y. Wiaux are with the Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. G. Puy is also with the Institute of the Physics of Biological Systems, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. Y. Wiaux is also with the Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, and with the Department of Radiology and Medical Informatics, University of Geneva (UniGE), CH-1211 Geneva, Switzerland. E-mail: gilles.puy@epfl.ch; pierre.vanderghenst@epfl.ch; yves.wiaux@epfl.ch - Address: EPFL STI IEL LTS2 - ELE 227 - Station 11 - CH-1015 Lausanne

This work is supported in part by the Center for Biomedical Imaging of the Geneva and Lausanne Universities, EPFL, and the Leenaards and Louis-Jeantet foundations, also by the Swiss National Science Foundation under grant PP00P2-123438, and by the EPFL-Merck Serono Alliance award.

<sup>1</sup> $\|\alpha\|_1 = \sum_{1 \leq i \leq N} |\alpha_i|$  ( $|\cdot|$  denotes the complex magnitude).

In this setting, common strategies focus on uniform random selection of the indices  $l_1, \dots, l_m$ . For signals sparse in the Dirac basis, a uniform random selection of Fourier basis vectors represents the best sampling strategy. Indeed, the Dirac and Fourier basis are optimally incoherent. Natural signals are however rather sparse in multi-scale bases, e.g. wavelet bases, not optimally incoherent with the Fourier basis. Many measurements are thus needed to reconstruct such signals accurately. This is for example the case in magnetic resonance imaging (MRI). To reduce the number of measurements, the authors in [3] rely on the fact that the energy of MRI signals is essentially concentrated at low frequencies. They thus propose to select Fourier basis vectors according to a variable density sampling profile selecting more low frequencies than high frequencies. This approach was shown to drastically enhance the quality of the reconstructed signals. This method is however essentially empirical and the reconstruction quality depends on the shape of the sampling profile used. Let us also mention that a line of justification for VDS was proposed in terms of the variable sparsity of the signals of interest as a function of scale in a wavelet sparsity basis [1], [4].

In this letter, we study VDS in the theoretical framework of compressed sensing. In Section II, we describe the latest compressed sensing results for sparse signals probed in bounded orthonormal system, and explain how they encompass variable density sampling procedures. In Section III, we introduce a minimization problem for the coherence between the sparsity and sensing bases, whose solution provides an optimized sampling profile. This minimization problem is solved with the use of convex optimization algorithms. We also propose a further refinement of our technique when prior information is available on the signal support  $S$ . In Section IV, we illustrate the effectiveness of the method through numerical simulations. We also provide a comparison of the Fourier VDS profile in the presence of prior information and corresponding reconstruction qualities, with the state-of-the-art VDS approaches used in MRI. Finally, we conclude in Section V.

## II. VARIABLE DENSITY SAMPLING

In the setting presented in Section I, the compressed sensing theory demonstrates that if the sampling indices  $l_1, \dots, l_m$  are chosen randomly and independently according to a discrete probability measure  $P$  defined on  $\{1, \dots, N\}$ , then a small number  $m \ll N$  of random measurements are sufficient for an exact reconstruction of  $\alpha$  [2].

**Theorem 1** (Theorem 4.2, [2]). *Let  $A = \Phi^\dagger \Psi \in \mathbb{C}^{N \times N}$ , and  $\alpha \in \mathbb{C}^N$  be a  $s$ -sparse vector such that<sup>2</sup>  $\text{sgn}(\alpha_S) \in \mathbb{C}^s$  is a random Steinhaus sequence. Assume that the sampling indices*

<sup>2</sup> $\text{sgn}(\alpha_S) \in \mathbb{C}^s$  is the  $s$ -dimensional vector with entries  $\alpha_i / |\alpha_i|, \forall i \in S$ .

$\Omega = \{l_1, \dots, l_m\}$  are selected randomly and independently according to a discrete probability measure  $P$  defined on  $\{1, \dots, N\}$ . Let  $\mathbf{y} = \mathbf{A}_\Omega \boldsymbol{\alpha} \in \mathbb{C}^m$  and define

$$\mu(P) = \frac{1}{N^{1/2}} \max_{1 \leq i, j \leq N} \frac{|\langle \phi_i, \psi_j \rangle|}{P^{1/2}(i)}. \quad (3)$$

For a universal constant  $C > 0$ , if

$$m \geq CN\mu^2(P)s \log^2(6N/\varepsilon), \quad (4)$$

then  $\boldsymbol{\alpha}$  is the unique minimizer of the  $\ell_1$ -minimization problem (2) with probability at least  $1 - \varepsilon$ .

In the above theorem, the parameter  $\mu(P)$  stands for the mutual coherence between the measurement basis  $\Phi$  and the sparsity basis  $\Psi$ . This value depends on the probability measure  $P$  and satisfies  $\mu(P) \geq N^{-1/2}$  [2]. The smaller the mutual coherence the smaller the required number of measurements for exact recovery.

Let us highlight that with the selection procedure described in Theorem 1, the number of measurements is exactly  $m$  but one measurement vector might be selected more than once. This characteristic is not always suitable in practical applications, such as MRI, particularly in a VDS configuration. Indeed, a sensing basis vector  $\phi_i$ , whose associated probability of selection  $P(i)$  is high, will be selected multiple times thus reducing the quantity of information probed. To avoid this phenomenon, we propose another selection process.

In the remainder, the sampling indices are selected according to an admissible sampling profile for  $m$  measurements.

**Definition 1** (Admissible sampling profile). A vector  $\mathbf{p} = (p_j)_{1 \leq j \leq N} \in \mathbb{R}^N$  is an admissible sampling profile for a number  $m$  of measurements if  $p_j \in (0, 1]$  for all  $1 \leq j \leq N$ , and  $\|\mathbf{p}\|_1 = m$ . The set of all admissible sampling profiles for a number  $m$  of measurements is denoted  $\mathcal{P}(m)$ .

Let  $\mathbf{p} \in \mathcal{P}(m)$  be an admissible sampling profile, the sampling indices are selected by generating a sequence  $(\delta_1, \dots, \delta_N) \in \mathbb{R}^N$  of independent Bernoulli random variables taking value 0 or 1 and such that  $\delta_i$  is equal to 1 with probability  $p_i$  for all  $1 \leq i \leq N$ . The set of selected indices is then defined as  $\Omega = \{l : \delta_l = 1\}$ . With the proposed sampling strategy, one measurement vector can be selected only once. The constraint that  $\|\mathbf{p}\|_1 = m$  imposes that the number of measurements is  $m$  on average over realizations of a sequence  $(\delta_1, \dots, \delta_N)$ . Note that for  $N \gg 1$ , the variability of the number of measurements is negligible.

As suggested in [2], one can actually show that the recovery condition (4) still holds with the coherence

$$\mu(\mathbf{p}) = \left(\frac{m}{N}\right)^{1/2} \max_{1 \leq i, j \leq N} \frac{|\langle \phi_i, \psi_j \rangle|}{p_i^{1/2}}. \quad (5)$$

The required elements of proof are provided in Appendix A.

### III. SAMPLING PROFILE OPTIMIZATION

Let us assume that the number of measurements  $m$  is fixed. In order to recover the highest sparsity  $s$  possible, Theorem 1 shows that we should use the sampling profile  $\mathbf{p} \in \mathcal{P}(m)$

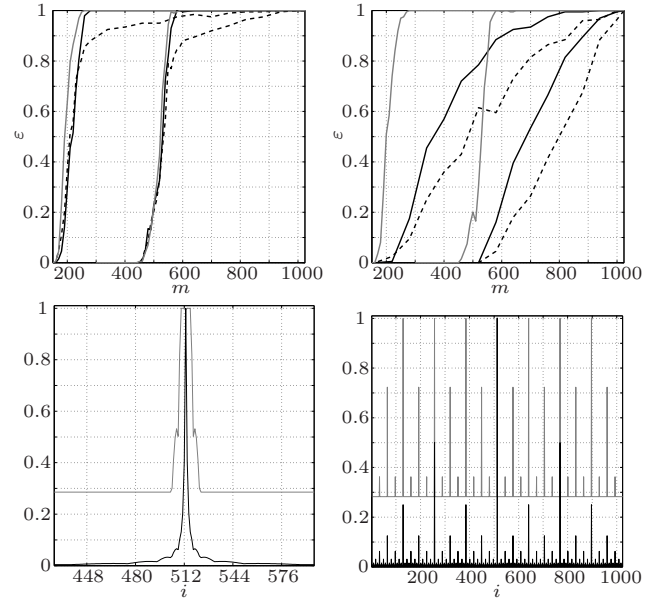


Fig. 1. Top panels: probability of recovery  $\epsilon$  of  $s$ -sparse signals in the Haar wavelet basis ( $N = 1024$ ) as a function of the number of measurements  $m$  in the Fourier basis (left panel) and in the Hadamard basis (right panel). The dark dashed, dark continuous, and light continuous curves show the probability of recovery with a uniform sampling, an optimized variable density sampling, and the spread spectrum technique respectively. Curves on the left correspond to  $s = 50$  and those on the right to  $s = 200$ . Bottom panels: light curves show the optimized sampling profile for  $m = 300$  obtained with a sampling in the Fourier basis (left panel) and in the Hadamard basis (right panel). Dark curves show the values  $\max_{1 \leq j \leq N} |\langle \phi_i, \psi_j \rangle|^2$  for all  $1 \leq i \leq N$ .

minimizing the mutual coherence  $\mu(\mathbf{p})$ . Therefore, we propose to solve the following optimization problem

$$(\hat{\mathbf{p}}, \hat{\mathbf{q}}) = \arg \min_{(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^{N \times 2}} \|\mathbf{B}\mathbf{q}\|_\infty + \lambda \|\mathbf{p} \cdot \mathbf{q} - \mathbf{1}\|_2^2 \text{ s.t. } \mathbf{p} \in \mathcal{K}_\tau, \quad (6)$$

where  $\lambda \in [0, +\infty)$ ,  $\tau \in (0, 1]$ ,  $\mathcal{K}_\tau = \{\mathbf{p} \in [\tau, 1]^N : \|\mathbf{p}\|_1 \leq m\}$ ,  $\mathbf{1} \in \mathbb{R}^N$  is the vector with all its entries equals to 1,  $\mathbf{p} \cdot \mathbf{q}$  is the entry-by-entry multiplication between the vector  $\mathbf{p}$  and  $\mathbf{q}$ ,  $\|\cdot\|_2$  and  $\|\cdot\|_\infty$  are respectively the  $\ell_2$ -norm and  $\ell_\infty$ -norm<sup>3</sup>, and  $\mathbf{B} \in \mathbb{C}^{N \times N}$  is the diagonal matrix with entries  $\max_{1 \leq j \leq N} |\langle \phi_i, \psi_j \rangle|^2$  on the diagonal,  $1 \leq i \leq N$ .

In the above problem, the term  $\|\mathbf{p} \cdot \mathbf{q} - \mathbf{1}\|_2^2$  ensures that  $p_i \simeq 1/q_i$  for all  $1 \leq i \leq N$ . The higher the value of the parameter  $\lambda$  the further this constraint is enforced. In the limit where  $\mathbf{p} \cdot \mathbf{q} = \mathbf{1}$ , we have  $\|\mathbf{B}\mathbf{q}\|_\infty = \mu^2(\mathbf{p})$  confirming that problem (6) seeks to minimize the mutual coherence. Note that the minimization problem imposes that  $\mathbf{p}$  belongs to the set  $\mathcal{K}_\tau$  which is different from the set  $\mathcal{P}(m)$ . Consequently, we do not have necessarily  $\|\mathbf{p}\|_1 = m$ . However, we note that in practice the constraint  $\|\mathbf{p}\|_1 \leq m$  is always saturated for high enough values of  $\lambda$ .

To solve problem (6), we adopt the following procedure:

- 1: Set  $t = 0$  and  $\hat{\mathbf{p}}^{(0)} = (m/N)_{1 \leq j \leq m}$ ;
- 2: **repeat**
- 3:  $\hat{\mathbf{q}}^{(t)} \leftarrow \arg \min_{\mathbf{q} \in \mathbb{R}^N} \|\mathbf{B}\mathbf{q}\|_\infty + \lambda \|\hat{\mathbf{p}}^{(t)} \cdot \mathbf{q} - \mathbf{1}\|_2^2$ ;
- 4:  $\hat{\mathbf{p}}^{(t+1)} \leftarrow \arg \min_{\mathbf{p} \in \mathbb{R}^N} \|\mathbf{p} \cdot \hat{\mathbf{q}}^{(t)} - \mathbf{1}\|_2^2$  s.t.  $\mathbf{p} \in \mathcal{K}_\tau$ ;
- 5:  $t \leftarrow t + 1$ ;
- 6: **until** convergence

<sup>3</sup> $\|\mathbf{x}\|_2^2 = \sum_{1 \leq j \leq N} |x_j|^2$  and  $\|\mathbf{x}\|_\infty = \max_{1 \leq j \leq N} |x_j|$ .

Subproblems at step 3) and 4) are convex problems. The subproblem at step 3) is solved iteratively using a forward-backward algorithm and the one at step 4) thanks to a parallel proximal algorithm [5]. Both algorithms require the computation of simple proximity operators. The computation of the one corresponding to  $\|B \cdot\|_\infty$  essentially reduces to a projection onto an  $\ell_1$ -ball (see Appendix B). This projection, as well as the one onto the  $\ell_1$ -ball of radius  $m$ , can be computed using the method<sup>4</sup> presented in [6]. Note that for both subproblems, the computational complexity at each iteration is essentially driven by these projections for which the method in [6] has a worst-case complexity of  $\mathcal{O}(N \log N)$ . For  $N = 1024$ , as in the forthcoming experiments, the overall algorithm converges in at most a few seconds. Our procedure therefore easily scales to larger  $N$ .

When prior information is available on the signal support  $S$ , we can refine our technique to find a sampling profile adapted to this support. Indeed, if the signal support  $S$  is known in advance then Theorem 1 applies with the coherence

$$\mu(p, S) = \left( \frac{m}{sN} \max_{1 \leq i \leq N} \frac{\sum_{j \in S} |\langle \phi_i, \psi_j \rangle|^2}{p_i} \right)^{1/2}. \quad (7)$$

We let the reader refer to Appendix A and equation (8) for more details. An optimized sampling profile associated with the set  $S$  can thus be obtained by substituting the diagonal matrix  $C \in \mathbb{R}^{N \times N}$  with entries  $s^{-1} \sum_{j \in S} |\langle \phi_i, \psi_j \rangle|^2$  on the diagonal,  $1 \leq i \leq N$ , for the matrix B in problem (6).

#### IV. EXPERIMENTS

In order to evaluate the proposed method in a general setting, we conduct two experiments. For the first one, we choose the Haar wavelet basis as the sparsity basis  $\Psi$  and the Fourier basis as the sensing basis  $\Phi$ . We generate complex  $s$ -sparse signals of size  $N = 1024$  with  $s \in \{50, 200\}$ . The positions of the non-zero coefficients are chosen uniformly at random in  $\{1, \dots, N\}$ , their phases are set by generating a Steinhaus sequence, and their amplitudes follows a uniform distribution over  $[0, 1]$ . The signals are then probed according to relation (1) and reconstructed from different number of measurements  $m$  by solving the  $\ell_1$ -minimization problem (2) with the SPGL1 toolbox [6]. For each value of  $m$ , the selected sensing basis vectors are chosen using the method described in Section II using either a uniform density profile or the profile  $\hat{p}$  obtained by solving problem (6) with  $\lambda = 0.05$ . Each time, the probability of recovery<sup>5</sup> is computed over 200 simulations. For the second experiment, the same setting is used but with the Hadamard basis as the sensing basis  $\Phi$ .

In order to evaluate our method when prior information is available on the support  $S$ , we perform a simplified MRI experiment. In this perspective, an *in vivo* brain image of size  $256 \times 256$  was acquired on a 7 Tesla scanner (Siemens, Erlangen, Germany). As suggested in [3], we consider a Daubechies-4 wavelet basis as sparsity basis and decompose

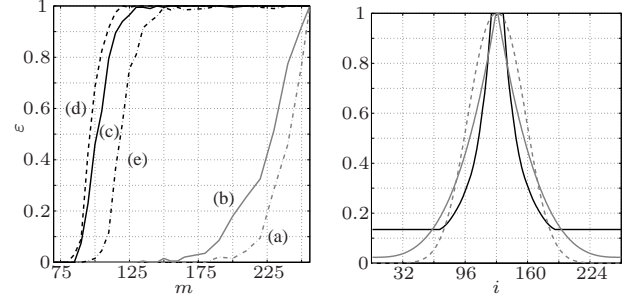


Fig. 2. Left panel: probability of recovery  $\epsilon$  of a MRI signal as a function of the number of measurements  $m$ . The light dot-dashed, light continuous, dark dot-dashed, dark continuous and dark dashed curves show, respectively, the probability of recovery obtained with: a uniform profile (a); the optimized profile obtained with the matrix B (b); the sampling profile proposed in [4] (e); the optimized profile obtained with the matrix C (c); the typical MRI profile used in [3] (d). Right panel: Optimized sampling profile obtained with the matrix C for  $m = 70$  (dark continuous curve) in comparison with the one used in MRI (light continuous curve) and the one proposed in [4] (light dashed curve).

each line<sup>6</sup> of the brain image into this basis. The resulting vectors are then hard-thresholded at  $s = 50$ . All vectors but one are seen as a data set providing prior information on the support  $S$  of typical MRI signals. The average of the values  $\sum_{j \in S} |\langle \phi_i, \psi_j \rangle|^2$ , for each  $i$  in  $\{1, \dots, N\}$ , serves to create the matrix C. The remaining signal, not considered in the data set, is considered as the signal under scrutiny, probed according to relation (1) and reconstructed from different number of measurements  $m$  by solving the  $\ell_1$ -minimization problem. For each value of  $m$ , the selected Fourier basis vectors are chosen using the method described in Section II with: a uniform density profile (a); the optimized sampling profile  $\hat{p}$  obtained with  $\lambda = 0.05$  and the matrix B (b) or with the matrix C (c); a typical sampling profile used in MRI (d) [3]; the sampling profile<sup>7</sup> proposed in [4] (e).

Figure 1 shows the probability of recovery  $\epsilon$  of  $s$ -sparse signals as a function of the number of measurements for the first two experiments. The probability of recovery obtained with the spread spectrum technique is also presented [7], [8]. Note that this technique, as for random Gaussian matrices, was proved to be universal, i.e., the number of measurements for the recovery of sparse signals is reduced to its minimum independently of the sparsity basis. One can note that the probability of recovery with the optimized sampling is always better than with the uniform sampling. With a sampling in the Fourier basis, one can also note that the recovery becomes almost optimal. Indeed, the number of measurements needed to reach a probability 1 of recovery is almost the same with the spread spectrum technique and with an optimized profile. These results confirm our theoretical predictions and illustrate the efficiency of variable density sampling.

As illustration, Figure 1 also shows optimized sampling profiles obtained for the two sensing bases and  $m = 300$  as well as the corresponding values of the diagonal entries of the matrix B. One can note that the shapes of the sampling

<sup>4</sup>Code available at <http://www.cs.ubc.ca/labs/scl/spgl1>

<sup>5</sup>Perfect recovery is considered if the  $\ell_2$ -norm between the original signal  $x$  and the reconstructed signal  $x^*$  satisfies:  $\|x - x^*\|_2 \leq 10^{-3} \|x\|_2$ .

<sup>6</sup>Lines without any signal (background) are withdrawn. After this operation, 151 lines are left.

<sup>7</sup>Note that the intrinsic parameters the sampling profiles (d) and (e) are manually chosen to obtained the best reconstructions.



profiles are highly correlated to the values in the matrix B.

Figure 2 shows the probability of recovery  $\varepsilon$  of the MRI signal as a function of the number of measurements. One can note that with the uniform sampling (a), the signal is recovered with probability 1 only when  $m = N$ . The results are slightly improved with the optimized profile (b) obtained with the matrix B. The sampling profiles (c), (d), and (e) drastically enhance the performance. Our optimized profile (c) obtained with the matrix C performs better than the profile (e) and similarly to the profile (d) typically used in MRI. These results provide a theoretical underpinning to VDS procedures used in MRI. It also shows that the refinement proposed for our technique can drastically enhance the performance of compressed sensing in practical applications.

For illustration, Figure 2 also shows the sampling profiles (c), (d), and (e) for  $m = 70$ . One can notice that the profiles (c) and (d) are very similar to each other. This explains again the effectiveness of VDS profiles commonly used in MRI.

## V. CONCLUSION

In the aim of optimizing variable density sampling profiles in the context of compressed sensing, we have introduced a minimization problem for the coherence between the sparsity and sensing bases. This problem is solved with the use of convex optimization algorithms. We have also discussed a refinement of our technique when prior information is available on the signal support in the sparsity basis. The effectiveness of the method is confirmed by numerical experiments. In particular, for signals sparse in a wavelet basis and probed in the Fourier domain, simulations show that our technique leads to optimal recovery. Indeed our technique gives similar probabilities of recovery as the spread spectrum method recently proved to be optimal. Our results also provide a theoretical underpinning to VDS procedures used in MRI.

## APPENDIX A

The proof of this theorem follows exactly the method used to prove Theorem 4.2. in [2]. The only difference resides in the estimate of the singular values of the operator  $A_{\Omega S}^\dagger A_{\Omega S}$  (see Theorem 7.3, [2]), where  $A_{\Omega S} \in \mathbb{C}^{m \times s}$  is the restriction of the matrix  $A_\Omega$  to the columns indexed by  $S$ .

**Lemma 1.** *Let  $A = \Phi^\dagger \Psi \in \mathbb{C}^{N \times N}$ ,  $\mathbf{p} = \{p_i\}_{1 \leq i \leq N} \in \mathcal{P}(m)$ ,  $\delta \in (0, 1/2]$ , and define  $P \in \mathbb{R}^{N \times N}$  the diagonal matrix with entries  $p_i^{1/2}$  on the diagonal,  $1 \leq i \leq N$ . Assume that the  $m$  measurement vectors are selected according to  $\mathbf{p}$  and suppose that  $s \geq 2$ . For a universal constant  $C > 0$ , the normalized matrix  $\tilde{A} = P^{-1}A$  satisfies  $\|\tilde{A}_{\Omega S}^\dagger \tilde{A}_{\Omega S} - I\| \leq \delta$  with probability at least  $1 - 2^{3/4}s \exp\left[-\frac{m\delta^2}{CN\mu^2(\mathbf{p})s}\right]$ .*

*Proof:* Let us denote  $Y = \tilde{A}_{\Omega S}^\dagger \tilde{A}_{\Omega S} - I = \sum_{i=1}^N \delta_i \tilde{\mathbf{a}}_i^\dagger \tilde{\mathbf{a}}_i - I \in \mathbb{C}^{s \times s}$  where  $\tilde{\mathbf{a}}_i \in \mathbb{C}^{1 \times s}$  is the  $i^{\text{th}}$  row of  $\tilde{A}_S$ . The proof starts by noticing that  $\mathbb{E}[Y] = \sum_{i=1}^N p_i \tilde{\mathbf{a}}_i^\dagger \tilde{\mathbf{a}}_i - I = \sum_{i=1}^N \mathbf{a}_i^\dagger \mathbf{a}_i - I = A_S^\dagger A_S - I = 0$ . We can thus continue with the use of a symmetrization technique to bound the expected value of the norm of  $Y$  (see Lemma 6.7 in [2], or proof of Theorem 3.1 in [1]). Let  $(\epsilon_1, \dots, \epsilon_N)$  be a

Rademacher sequence independent of  $(\delta_1, \dots, \delta_N)$  and  $p \geq 2$ , then  $\mathbb{E}\|Y\|^p \leq 2^p \mathbb{E}\|\sum_{1 \leq i \leq N} \epsilon_i \delta_i \tilde{\mathbf{a}}_i^\dagger \tilde{\mathbf{a}}_i\|^p$ . Noticing that  $\tilde{A}_{\Omega S}$  has at most rank  $s$ , using Fubini's theorem, Rudelson's lemma (see Lemma 6.18, [2]) conditional on  $(\delta_1, \dots, \delta_N)$ , and the Cauchy Schwarz inequality yields

$$\mathbb{E}\|Y\|^p \leq 2^{3/4+p} s \left(\frac{p}{e}\right)^{p/2} \sqrt{\mathbb{E}\|\tilde{A}_{\Omega S}^\dagger \tilde{A}_{\Omega S}\|^p \mathbb{E}\left[\max_{i: \delta_i=1} \|\tilde{\mathbf{a}}_i\|_2^{2p}\right]}.$$

The previous equation is identical to equation (7.6) in the proof of Theorem 7.3 in [2]. We can follow the same remaining steps of this proof to terminate ours. We still need however to provide a bound on  $\max_{i: \delta_i=1} \|\tilde{\mathbf{a}}_i\|_2^2$ . If the support  $S$  is fixed and known in advance, we have

$$\max_{i: \delta_i=1} \|\tilde{\mathbf{a}}_i\|_2^2 \leq \max_{1 \leq i \leq N} \frac{\sum_{j \in S} |\langle \phi_i, \psi_j \rangle|^2}{p_i} = \frac{sN}{m} \mu(\mathbf{p}, S)^2. \quad (8)$$

In the general case where  $S$  is unknown, we can write

$$\max_{i: \delta_i=1} \|\tilde{\mathbf{a}}_i\|_2^2 \leq s \max_{1 \leq i, j \leq N} \frac{|\langle \phi_i, \psi_j \rangle|^2}{p_i} = \frac{sN}{m} \mu(\mathbf{p})^2. \quad (9)$$

## APPENDIX B

The proximity operator of  $\gamma\|B \cdot\|_\infty$ ,  $\gamma > 0$ , is the unique solution of  $\text{prox}_{\gamma\|B \cdot\|_\infty}(\mathbf{q}) = \arg \min_{\mathbf{x} \in \mathbb{R}^N} 1/2\|\mathbf{q} - \mathbf{x}\|_2^2 + \gamma\|B \mathbf{x}\|_\infty$ .

**Proposition 1.** *For any  $\mathbf{q} \in \mathbb{R}^N$ , with  $B \in \mathbb{R}^{N \times N}$  defined as above, we have  $\text{prox}_{\gamma\|B \cdot\|_\infty}(\mathbf{q}) = \mathbf{q} - \gamma \text{proj}_{\mathcal{C}}(\mathbf{q}/\gamma)$ , where  $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^N : \|B^{-1}\mathbf{x}\|_1 \leq 1\}$  and  $\text{proj}_{\mathcal{C}}$  denotes the projection onto the set  $\mathcal{C}$ .*

*Proof:* From Theorem 14.3 in [5], we have  $\mathbf{q} = \text{prox}_{\gamma\|B \cdot\|_\infty}(\mathbf{q}) + \gamma \text{prox}_{\gamma^{-1}\|B \cdot\|_\infty}^*(\mathbf{q}/\gamma)$ , for all  $\mathbf{q} \in \mathbb{R}^N$ . In the previous relation,  $\|B \cdot\|_\infty^*$  denotes the Fenchel conjugate of  $\|B \cdot\|_\infty$ . As  $B$  is a bijection (it is a diagonal matrix with strictly positive entries), one can show that  $(\|B \cdot\|_\infty)^* = \iota_{\mathcal{C}}(\cdot)$  where  $\iota_{\mathcal{C}}$  denotes the indicator function of the set  $\mathcal{C}$  (Proposition 13.20, [5]). Finally, we have  $\text{prox}_{\gamma^{-1}\|B \cdot\|_\infty}^* = \text{prox}_{\gamma^{-1}\iota_{\mathcal{C}}}(\cdot) = \text{proj}_{\mathcal{C}}$ . Combining the last result with the first relation terminates the proof. ■

## REFERENCES

- [1] E. J. Candès and J. Romberg, "Sparsity and Incoherence in Compressive Sampling" *Inverse Problems*, vol. 23, pp. 969-985, 2007.
- [2] H. Rauhut, "Compressive Sensing and Structured Random Matrices," *Radon Series Comp. Appl. Math.*, vol. 9, pp. 1-92, 2010.
- [3] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182-1195, 2007.
- [4] Z. Wang and G. R. Arce, "Variable Density Compressed Image Sampling," *IEEE Trans. Image Process.*, vol. 19, pp. 264-270, 2010.
- [5] H. H. Bauschke and P. L. Combettes, "Convex Analysis and Monotone Operator Theory in Hilbert Spaces," Springer-Verlag, New York, 2011.
- [6] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, pp. 890-912, 2008.
- [7] G. Puy, P. Vandergheynst, R. Gribonval, and Y. Wiaux, "Spread Spectrum for Universal Compressive Sampling," *4th Workshop on Signal Processing with Adaptive Sparse Structured Representations*, Edinburgh, pp. 50, 2011.
- [8] G. Puy, P. Vandergheynst, R. Gribonval, and Y. Wiaux, "Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques," *EURASIP Journal on Advances in Signal Processing*, submitted, 2011.